

# Bayesian Entropic Inverse Theory Approach to Implied Option Pricing with Noisy Data

Igor Halperin

BFM Financial Research

February 1, 2008

e-mail: *i\_halperin@hotmail.com*

## Abstract:

A popular approach to nonparametric option pricing is the Minimum Cross Entropy (MCE) method based on minimization of the relative Kullback-Leibler entropy of the price density distribution and a given reference density, with observable option prices serving as constraints. When market prices are noisy, the MCE method tends to overfit the data and often becomes unstable. We propose a non-parametric option pricing method whose input are noisy market prices of arbitrary number of European options with arbitrary maturities. Implied transition densities are calculated using the Bayesian inverse theory with entropic priors, with a reference density which may be estimated by the algorithm itself. In the limit of zero noise, our approach is shown to reduce to the canonical MCE method generalized to a multi-period case. The method can be used for a non-parametric pricing of American/Bermudan options with a possible weak path dependence.

---

Opinions expressed in this paper are those of the author, and not necessarily of his employer. I want to thank E. Berger for useful discussions and reading the manuscript. All remaining errors are mine, of course.

# 1 Introduction

The linear inverse theory deals with the problem of reconstruction of a signal  $\{f_i\}$  out of data (images)  $D = \{d_1, \dots, d_m\}$  which are obtained from a signal by a linear operator  $K$ , and may be corrupted with a noise  $\varepsilon$ :

$$d_i = \sum_{j=1}^n K_{ij} f_j + \varepsilon_i \quad i = 1, \dots, m \quad (1)$$

Since the number of available data  $m$  is typically much smaller than the dimension  $n$  of the probability space, inverse problems are highly underdetermined and ill-posed. Additional considerations, such as a prior knowledge and/or guiding regularization principles, become of a critical importance for solving such a problem. Inverse problems are widely encountered in fields as diverse as image processing, geophysics, speech recognition, computational linguistics, molecular biology, econometrics etc. Much progress in developing reliable and robust methods of inversion was made in these areas over last decades, with mainstream efforts concentrated on Bayesian and Maximum Entropy (MaxEnt) or Minimum Cross Entropy (MCE) methods, see e.g. [1, 2, 3, 4].

A popular paradigm in option pricing theory is the nonparametric option pricing which is based on inferring a (risk-neutral) price probability density out of market prices of liquid assets, and then calculating prices of other, typically less liquid instruments as convolutions of payoff functions with the inferred distribution. Depending on the data, a solution may not exist (that would signal arbitrage opportunities) or, more often, there may be multiple probability measures consistent with the same data (this means that the market is incomplete), thus the question of choosing the “right” solution arises. The problem of obtaining asset price distributions out of option prices is thus a typical (linear) inverse problem, which can be addressed using well established methods of the inversion theory.

One popular method to find implied densities is to use the MCE inversion method as advocated and developed by Jaynes [1], without invoking Bayesian arguments. This approach was initiated in the financial literature by Gulko [5], Buchen and Kelly [6], and Stutzer [7]. Within this method, the asset price distribution  $p(x)$  is found by requiring that it should minimize the relative entropy of  $p(x)$  and a given reference distribution  $q(x)$ , while reproducing the market prices of options in a chosen calibrating set (with all options having the same maturity). The relative entropy of two distributions  $p(x)$  and  $q(x)$  is given by the so-called Kullback-Leibler (KL) distance [8]

$$D[p(x)||q(x)] = \int dx p(x) \log \frac{p(x)}{q(x)} \quad (2)$$

As  $\log(1/x)$  is a convex function, Jensen’s inequality

$$\int dy p(y) g[q(x)] \geq g \left( \int dy p(y) q(y) \right) \quad (3)$$

valid for any convex function  $g$  and probability  $p(y)$  such that  $\int dy p(y) = 1$ , shows that  $D[p(x)||q(x)] \geq 0$  for any distributions  $p(x), q(x)$ .

The reference density  $q(x)$  in (2) accumulates our prior knowledge and/or expectation for the answer. There exist several possibilities to choose such a density. One of them is to take  $q(x)$  to be the historical asset price distribution, as suggested by Stutzer [7]. On the other hand, if no historical information is available or if one believes that the risk-neutral distribution should be very different from the historical one, one may take a lognormal distribution or even a uniform reference distribution  $q(x) = \text{const}$ . In the latter case, the KL distance becomes the (minus) absolute entropy of distribution  $p(x)$ . The principle of maximizing the absolute entropy of a distribution (MaxEnt) is thus but a particular case of a more general principle of minimizing the relative KL entropy (MCE).

With a uniform reference density  $q(x) = \text{const}$ , the MaxEnt method gives the least informative and least prejudiced asset price distribution compatible with the market data. As was stressed by Gulko (who considered the case  $q(x) = \text{const}$  only), the MaxEnt principle thus formalizes and quantifies the Efficient Market Hypothesis in the form of a static characterization of market beliefs about future prices: the distribution must be such that it maximizes uncertainty about the future price movements. By the meaning of entropy [8], this means that the distribution should be of a highest possible absolute entropy, i.e. the lowest possible KL distance to a uniform distribution. The argument thus obviously generalizes to the case of a non-uniform density  $q(x)$  expressing the market belief about the future: the most probable distribution  $p(x)$  should minimize the KL distance to the reference distribution  $q(x)$ , while being consistent with the market data. The MCE approach thus provides a consistent and theoretically sound method of “completing the market”, i.e. choosing the “best”, maximally unbiased risk-neutral probability measure if several possibilities compatible with the data are possible.

In the early works [5, 6, 7], the MCE method was applied to a simplest one-period model, with all options in the calibrating set having the same maturity. Furthermore, no noise was supposed to be present in the data. Later, the method was extended to a multi-period setting by Avellaneda and co-workers [9] using methods of stochastic optimal control theory, again without allowing for a noise in the data. For other recent applications of the canonical MCE approach, see Derman and Zou [10] and Stutzer [11].

In reality, noise is always present in option prices, and can have a number of sources including non-synchronicity of trading, straight-out reporting errors, different trader perspectives of the market, etc. As was found by Buchen and Kelly [6] with Monte Carlo simulation, the MCE method often becomes unstable even for a modestly low level of noise in the data, leading sometimes to a fake multi-nodal form of an implied distribution. This behavior is traced back to the construction of the canonical MCE method (see below in Sect.3) which enforces exact matching of the market prices and thus tends to overfit the data. These problems have led Buchen and Kelly to the conclusion that the method should be abandoned altogether [13] in favor of the so-called geophysical inverse theory based on a completely different approach. However, a way to overcome these difficulties within a natural generalization of the canonical entropy approach has been known for a long time to information theory researchers: It was stressed by Jaynes as early as in 1982 [12] that when the data are noisy, the MaxEnt and MCE methods are unsuitable and should be substituted by the **Bayesian inverse theory in combination with the MCE principle**. There exists an extensive literature [12, 3] documenting both the need and the way to modify the canonical MCE method for dealing with noisy data.

The contribution of the present work is two-fold. First, we develop a Bayesian Inverse Theory approach to implied option pricing with noisy data, that naturally generalizes the canonical MCE method and reduces to the latter in the limit of zero noise. The only restriction we impose on the stochastic process for the asset price is the Markovian property. Beyond this, its nature (diffusion, jump-diffusion, etc.) is left unspecified. Second, we extend the method to the case of calibration to options of arbitrary maturities rather than one fixed maturity as in [5, 6, 7]. In contrast to Avellaneda *et.al* [9], we build a recursive procedure. We first calculate conditional transition probabilities with transition times being the adjacent maturities of options in the calibrating set. This is performed recursively starting from a shortest maturity. When this is done, we may, if needed, fill the model “in between” introducing intermediate time levels between the shortest and the longest maturities of the calibrating set. This may be necessary e.g. for pricing of a Bermudan option with exercise dates some (or all) of which do not belong in the given set of maturity dates. Our method allows one to introduce an arbitrary number of intermediate time levels and calculate corresponding conditional transition densities. The method involves two adjustable parameters related to modeling the noise and treating the reference density (see below) that may be either estimated from the data, or serve as a measure of the model risk. As the output of our method is a set of conditional transition densities, it may be considered as a generalization of the implied tree methodology of Rubinstein [14] and Derman and Kani [15] (see [16] for a review). In this class of models, closest to ours is presumably a semirecombining implied tree construction of Brown and Toft [17]. Similarly to the method of [17], we first find the conditional transition densities between different maturity dates in the calibrating set, and then calculate distributions for intermediate times of interest such as exercise dates for a Bermudan option. However, in contrast to [17], we do not discretize a stochastic process for the underlying, do not introduce further time levels and, most important, use a different approach to fix the transition densities, which also allows one to account for the noise in data. Computationally, our method is equivalent to performing a single MCE optimization for each maturity where data are available, and calculating four one-dimensional integrals for each additional time level. In addition, our algorithm is flexible in handling the reference density: at each step, it is generally given by a linear combination of a prior (e.g. historical) density and a conditional density of a previous step evolved to the current time (see Eq.(7) below). The attractive feature of allowing for a non-vanishing value of  $0 < \rho < 1$  is that it smoothes the transition density in the time direction.

Our presentation is organized as follows. In Sect.2 we introduce the basic recursive algorithm. The subsequent sections 3 to 7 describe a single time step of the algorithm. In Sect.3 we outline the MCE approach to the problem valid when no noise is present in the data. The Bayesian entropic inverse theory approach to the problem needed when data are noisy is presented in Sect.4. In Sect.5 we explain how the noise variance should be integrated out in our Bayesian setting. A resulting non-linear optimization problem with an objective function (which we call the effective Lagrangian using physics’ nomenclature) is formulated in Sect.6, and its convexity properties are studied. The actual minimization of the effective Lagrangian is performed in Sect.7 using a linearization trick that leads to a drastic reduction in the computational cost. Sect.8 explains how to calculate transitions densities involving additional time levels. The final Sect.8 contains our conclusion.

## 2 Recursive Algorithm

We assume that we are given an ordered set of maturities  $\{T_1, \dots, T_n\}$ . For each maturity  $T_i$ ,  $i = 1, \dots, n$ , there are  $N_i$  available European option prices. The option payoff functions are supposed to be linearly independent [6].

Our scheme can be introduced as follows. Let  $X_1, \dots, X_n$  be the asset prices at times  $\{T_1, \dots, T_n\}$ , respectively. We assume that a stochastic process governing the evolution of the asset price is Markovian. Time today is  $t = 0$ . Consider now the time moment  $T_{i-1}$  for some  $i$ . We ask what the asset price distribution should be at time  $T_i$ . Had we known the asset price  $X_{i-1} = x$  at time  $T_{i-1}$ , following the MCE approach of [5, 6, 7] we would minimize the KL entropy for the distribution  $p(X_i|X_{i-1} = x_{i-1})$ . Since at most we might know only the distribution  $p(x_{i-1})$  of the asset price at a future moment  $T_{i-1}$ , the best we can do is to require that KL distance should be minimized on average. In other words, we require minimization of the conditional KL entropy [8]

$$D[p(x_i|x_{i-1})||q(x_i|x_{i-1})] = \int dx p(x_{i-1}) \int dy p(x_i|x_{i-1}) \log \left( \frac{p(x_i|x_{i-1})}{q(x_i|x_{i-1})} \right) \quad (4)$$

with respect to  $p(x_i|x_{i-1})$  for each  $i = 1, \dots, n$ , subject to the condition of matching the observed market option prices. This suggests a forward recursive algorithm for solving the problem. (The matching of data can be done either in the form of absolute constraints or in the Bayesian framework via a likelihood function. As will be clear below, the latter preserves the forward structure of the relations between different distributions.)

It should be noted that this recursive algorithm is equivalent to a direct minimization of the KL entropy of the joint distribution  $p(X_1, \dots, X_n)$

$$D[p(x_1, \dots, x_n)||q(x_1, \dots, x_n)] = \int dx_1 \dots dx_n p(x_1, \dots, x_n) \log \frac{p(x_1, \dots, x_n)}{q(x_1, \dots, x_n)} \quad (5)$$

Indeed, using the chain rule for relative entropy [8], we obtain

$$\begin{aligned} D[p(x_1, \dots, x_n)||q(x_1, \dots, x_n)] &= \sum_{i=1}^n D[p(x_i|x_{i-1}, \dots, x_1)||q(x_i|x_{i-1}, \dots, x_1)] \\ &= \sum_{i=1}^n D[p(x_i|x_{i-1})||q(x_i|x_{i-1})] \end{aligned} \quad (6)$$

where on the last step we used the Markov property  $p(X_i|X_{i-1}, \dots, X_1) = p(X_i|X_{i-1})$  (and analogously for the reference distribution). As each term in the series (6) is non-negative, minimization of the whole expression amounts to a sequential minimization of each successive term in the series<sup>1</sup>. This is a special property of a Markov process which enables one to use a forward calculation to calculate probabilities<sup>2</sup>. The two schemes produce identical results, as can also be checked in a simple two-period setting. We note that it is the scheme of minimization of the joint KL distance (5) that was adopted by

---

<sup>1</sup>This statement is valid assuming that a minimum is unique. As will be argued below in Sect. 3, in our optimization problem the minimum, if exists, is unique.

<sup>2</sup>A similar forward procedure is used for implied trees in the approaches of [14, 15, 17].

Avellaneda *et.al* [9]. We prefer the former forward recursive scheme because it is easier to implement, and also because the conditional transition probabilities needed for pricing other securities are its direct outputs, while in the scheme adopted in [9] their calculation includes additional integrations. Last but not least, a generalization to the case of noisy data is quite straightforward in the recursive scheme but not in the method of [9].

Our recursive algorithm proceeds as follows. Suppose that the problem is solved up to time  $T_k$ , i.e. we know all conditional transition probabilities  $p(y, T_{i+1}|x, T_i)$  for  $i < k$  as well as the marginal density (conditional on the initial price)  $p(x) \equiv p(x, T_k|x_0, 0)$  where time today is  $t = 0$  and today's asset's value is  $x_0$ . All reference conditional transition densities are calculated as well. The recursion step is then performed as follows.

1. Set the transition times  $t \equiv T_k$ ,  $T \equiv T_{k+1}$ .
2. Calculate the conditional transition probability  $p(y|x) \equiv p(y, T|x, t)$  of the asset price distribution at time  $T$  conditional on its value at time  $t$ . The calculation is performed using the Bayesian inverse theory, and requires the knowledge of the marginal density  $p(x)$  at time  $t$ . It also involves the reference conditional transition density  $q(y|x)$  for this step.
3. Calculate the reference conditional transition density for the next step. To this end, we shift all time arguments (which enter the calculation through options payoff functions) in  $p(y, T_{k+1}|x, T_k)$  to the next time level  $(T_k, T_{k+1}) \rightarrow (T_{k+1}, T_{k+2})$ . Call this density  $q_s(y|x)$ . A general form for the reference density for the next step thus would be

$$q(y|x) = (1 - \rho) q_h(y|x) + \rho q_s(y|x) \quad (7)$$

where  $q_h(y|x)$  stands for the conditional density from time  $T_{k+1}$  to time  $T_{k+2}$  calculated with a prior model (or a historical price distribution), and  $0 \leq \rho \leq 1$  is a free parameter of the model. The limit  $\rho \rightarrow 1$  corresponds to maximum smoothing of a conditional transition density in the time direction<sup>3</sup>

4. Calculate a new marginal density

$$\int dx p(x)p(y|x) = \int dx p(x|x_0)p(y|x)$$

to be used to find the conditional transition density on the next step.

5. Increment  $k \rightarrow k + 1$ .
6. If  $k < n - 1$ , go to step 1.
7. If needed, calculate conditional transition densities involving additional time levels.

In what follows, Sects. 3 to 7 describe step 2 in the above scheme, while Sect.8 explains the calculation on step 7.

### 3 Canonical Minimum Cross Entropy Method

Assuming that the unconditional density  $p(x)$  at time  $t = T_k$  is known, we want to find the transition probability density  $p(y|x)$  for transition to the interval  $[y, y + dy]$  at time

---

<sup>3</sup>The choice of  $q_s(y|x)$  as a prior distribution is reasonable if the real process is stationary, which means that the conditional density  $p(y, T|x, t)$  depends only on the difference  $T - t$ . We can further neglect an implicit dependence of Lagrange multipliers on  $T - t$ , which is a good zero approximation so long as  $T_{k+2} - T_{k+1}$  is close (or equal) to  $T_{k+1} - T_k$ .

$T \equiv T_{k+1}$  conditional on being in state  $x$  at time  $T_i$ . We require that  $p(y|x)$  should be close to a prior density  $q(y|x)$ , i.e. should minimize the conditional KL entropy functional

$$D[p(y|x)||q(y|x)] = \int dx p(x) \int dy p(y|x) \log \frac{p(y|x)}{q(y|x)} \quad (8)$$

subject to all constraints imposed on  $p(y|x)$ . The first constraint is the normalization equation

$$\int dy p(y|x) = 1 \quad (9)$$

that should be imposed for all values of  $x$ . Further constraints are imposed by matching observed option prices. We consider a set of  $N \equiv N_{k+1}$  European options maturing at time  $T = T_{k+1}$ , with  $\{f_i(x, y)\}$  being the payoff functions. For the sake of generality, we retain the  $x$  argument in  $f_i(x, y)$ , which may account for a certain path dependence. More common, however, is calibration to European vanilla call and put options whose payoff functions  $f_i(x, y) = f_i(y)$  are

$$f_i(y) = D(T - t) \max(y - K_i, 0) \quad (\text{call}) \quad (10)$$

$$f_i(y) = D(T - t) \max(K_i - y, 0) \quad (\text{put}) \quad (11)$$

where  $\{K_i\}$  are strikes and  $D(T - t)$  are (deterministic) discount factors:  $D(T - t) = \exp[-r(T - t)]$ . An exact matching of market prices imposes the constraints

$$\int dx p(x) \int dy p(y|x) f_i(x, y) = C_i \quad , \quad i = 1, \dots, N \quad (12)$$

where  $C_i$  stands for a (noiseless) price of the  $i$ th option.

The problem of minimization of (8) subject to constraints (9) and (12) is a standard problem of constrained optimization that is solved by the Lagrange multiplier method. To satisfy the normalization constraint (9) for each value of  $x$ , one needs a Lagrange multiplier function  $\xi(x)p(x)$ . Here the multiplicative factor  $p(x)$  (known from the previous step) is introduced for a later convenience. Constraints (12) are imposed using a discrete set of  $N$  Lagrange multipliers  $\{\lambda_i\}$ . We thus arrive at the following Lagrangian function

$$\begin{aligned} L = & \int dx p(x) \int dy p(y|x) \log \frac{p(y|x)}{q(y|x)} - \int dx \xi(x) p(x) \left( \int dy p(y|x) - 1 \right) \\ & - \sum_i \lambda_i \left( \int dx p(x) \int dy p(y|x) f_i(x, y) - C_i \right) \end{aligned} \quad (13)$$

The Lagrangian (13) is minimized when its total variation vanishes, i.e.

$$\frac{\delta L}{\delta p(y|x)} = 0 \quad , \quad \frac{\delta L}{\delta \xi(x)} = 0 \quad , \quad \frac{\partial L}{\partial \lambda_i} = 0 \quad , \quad i = 1, \dots, N \quad (14)$$

The solution of the first two equations in (14) reads

$$p(y|x) = \frac{q(y|x)}{Z_\lambda(x)} \exp \left( \sum_i \lambda_i f_i(x, y) \right) \quad (15)$$

where

$$Z_\lambda(x) = \int dy q(y|x) \exp \left( \sum_i \lambda_i f_i(x, y) \right) \quad (16)$$

What remains is to eliminate the Lagrange multipliers  $\lambda_i$  by substituting the solution (15) in the last of Eqs.(14). This results in a set of nonlinear equations

$$\int \int p(x) \frac{q(y|x)}{Z_\lambda(x)} \exp \left( \sum_i \lambda_i f_i(x, y) \right) f_i(x, y) = C_i \quad , \quad i = 1, \dots, N \quad (17)$$

This problem can be solved e.g. by a multidimensional Newton-Raphson procedure [18]. As shown in [6], the Jacobian matrix in this problem is given by the covariance matrix of the constraint functions

$$J_{ij} = \text{cov} [f_i(x, y), f_j(x, y)] \quad (18)$$

and thus is positive definite for linearly independent constraints, which is a necessary condition for the Newton-Raphson method to be applicable. In addition,  $J_{ij}$  should be well-conditioned to ensure stability of the algorithm. As discussed in [6], the Jacobian may become ill-conditioned if market prices are noisy. In case a case, a numerical solution for the Lagrange multipliers may become unstable.

However, market prices are always noisy. As the classical MCE method enforces the exact matching of constraints (12), it tends to overfit the data. As a result, implied MCE probability distributions often become unstable and juggled. Intuitively, we might want to substitute the exact constraints (12) by approximate ones that would smooth out the noise in the data. It will be shown in the next sections that the Bayesian inverse theory with entropic priors provides just such a sort of modification of the MCE method.

## 4 Noisy Data: Bayesian Entropic Inverse Theory

In Bayesian probability theory [1], the posterior probability  $P(M|D, I)$  of a model (probability density, in our case)  $M$  given the data  $D$  and information  $I$  is determined by the Bayes formula

$$P(M|D, I) = P(M|I) \frac{P(D|M, I)}{P(D|I)} \quad (19)$$

where  $P(M|I)$  the prior probability of the model assigned before the data is observed, and  $P(D|M, I)$  is the likelihood function measuring probability to observe the data  $D$  given the model  $M$ . Finally, the so-called evidence  $P(D|I)$  is the total probability of observing the data given the information  $I$ . As it does not depend on the model  $M$  whose probability we want to calculate, it merely serves as an irrelevant normalization factor for the posterior probability  $P(M|D, I)$ .

To solve the inverse problem of finding the transition probabilities based on data and a prior knowledge, we apply the Maximum A-posteriori Probability (MAP) approach [1, 3]. According to this method, the most probable model is one that maximizes the posterior probability  $P(M|D, I)$  in (19) i.e. the product of the prior  $P(M|I)$  and the likelihood function  $P(D|M, I)$ . Note here the difference with the usual Maximum Likelihood method which requires fixing the prior in advance and optimizing the likelihood function alone.

On the contrary, the MAP method allows the prior to carry unknown parameters and/or functional dependence.

We start with assigning the prior probability  $P(M|I)$ . The following combinatorial argument [19] that extends one suggested by Jaynes [1], is of use here (see also [20] for a recent discussion of entropic priors). For simplicity, we will first consider marginal, not conditional probabilities, and discretize the state space  $X = \{1, 2, \dots, k\}$ . Let  $q_j$  be the probability of the  $j$ th element of  $X$ . We can imagine an urn with known proportions  $q_j$  of balls of type  $j$ , for  $j = 1, 2, \dots, k$ . If we randomly draw  $n$  balls from the urn, then the probability of observing  $n_i = p_i n$  balls of each class  $i$  is given by the multinomial distribution

$$P(p|q) = P(n_1, \dots, n_k|q) = \frac{n!}{n_1! n_2! \dots n_k!} q_1^{n_1} \dots q_k^{n_k} \quad (20)$$

which is the probability of seeing a  $p$ -distribution when sampling  $n$  times from a  $q$ -distribution. We assume that  $n \gg 1$ . Taking the logarithm of both sides of (20) and using the Stirling formula

$$\log m! = m \log m - m + o(m) \quad (21)$$

we obtain

$$P(p|q) \propto \exp \left( -n \sum_{j=1}^k p_j \log \frac{p_j}{q_j} \right) \quad (22)$$

Note that for the case of a uniform reference distribution  $q_j = q = \text{const}$ , this line of reasoning boils down to Jaynes' definition of the prior

$$P(p) \propto \exp \left( - \sum_{j=1}^k p_j \log p_j \right) \quad (23)$$

while the exact value of parameter  $n$  becomes irrelevant in this limit, as will be clear below. Extending the argument to the case of conditional probabilities, we write down the prior probability of distribution  $p(y|x)$  in our problem:

$$P[p||q] = \exp \left[ -h \int dx p(x) \int dy p(y|x) \log \frac{p(y|x)}{q(y|x)} \right] \prod_x \delta \left( \int dy p(y|x) - 1 \right) \quad (24)$$

where we added the  $\delta$ -function constraint for each value of  $x$  to ensure the correct normalization. Here  $h \gg 1$  is a parameter controlling the relative weight of the entropy term in the posterior probability.

Our next task is to compute the likelihood function  $P(D|M, I)$ . To this end, we assume that noise in prices is Gaussian<sup>4</sup>:

$$\int dx p(x) \int dy p(y|x) f_i(x, y) = C_i + \varepsilon_i \quad , \quad i = 1, \dots, N \quad (25)$$

---

<sup>4</sup>Note that the assumption of an additive noise in the prices is, strictly speaking, theoretically problematic, as potentially it allows for negative prices. However, assuming that a noise is not too large, a probability of this to occur will be negligible. Yet, presumably a more realistic approach would be to model an additive noise for logarithms of option prices. We will leave a study of such a model for a future work.

where

$$\mathbf{E}[\varepsilon_i] = 0 \quad , \quad \mathbf{E}[\varepsilon_i^2] = \sigma_i^2 \quad (26)$$

The likelihood function is therefore given by the following expression:

$$P[D|p, \sigma] = \frac{1}{(2\pi)^{N/2} \prod_i \sigma_i} \exp \left[ - \sum_i \frac{1}{2\sigma_i^2} \left( \int \int dx dy p(x) p(y|x) f_i(x, y) - C_i \right)^2 \right] \quad (27)$$

If the prior distribution for the noise variance is given by some function  $P(\sigma)$ , the full prior in our problem becomes <sup>5</sup>

$$P(M|I) = P(\sigma) P[p||q] \quad (28)$$

where  $P[p||q]$  is given by (24). As we are not interested in estimating the noise variances  $\sigma_i^2$ , they represent the so-called nuisance parameters (hyperparameters) that should be integrated out according to the Bayesian approach [1, 2]. Therefore, for the posterior probability we obtain

$$P[p|D] \propto P[p||q] \int_0^\infty \prod_i d\sigma_i P(\sigma) P[D|p, \sigma] \quad (29)$$

where  $P[p||q]$  and  $P[D|p, \sigma]$  are defined in (24) and (27), respectively. According to the MAP approach, this final expression should be maximized with respect to  $p(y|x)$  in order to find the most probable transition probability  $p(y|x)$  compatible with the data while remaining as close as possible to the reference density  $q(y|x)$ .

Eq.(29) allows one to readily see how the usual MCE method is reproduced in the limit of zero noise. This limit corresponds to knowledge of the prior for noise with certainty:

$$P(\sigma) = \prod_i \delta(\sigma_i - \sigma_i^{(0)}) \quad (30)$$

and taking the limit  $\sigma_i^{(0)} \rightarrow 0$  after the integral in (29) is calculated. This yields

$$P[p|D] \propto P[p||q] \prod_i \delta \left( \int \int dx dy p(x) p(y|x) f_i(x, y) - C_i \right) \quad (31)$$

Using the integral representation of the  $\delta$ -function

$$\delta(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega \exp(i\omega x) = \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} d\lambda \exp(-\lambda x) \quad (32)$$

we see that in the limit of zero noise the MAP approach exactly reproduces the MCE method, with the saddle point values of parameters  $\lambda$  in (32) for all  $\delta$ -functions in (31) being exactly equivalent to Lagrange multipliers in (13). It is easy to see that the exact value of parameter  $h$  becomes irrelevant in this limit, as it can be eliminated by a rescaling of Lagrange multipliers.

---

<sup>5</sup>A fully Bayesian approach would also require a specification of a prior for parameter  $h$  in (24). We will not attempt doing this, but will rather view  $h$  as a fixed parameter, and will study stability of our results with respect to variations of  $h$ .

## 5 Integrating out Noise Variance

In general, a noise in option prices does persist. To integrate it out according to the Bayesian theory, we should define a model for the prior distribution  $P(\sigma)$ . Given the prior, we then calculate the integral

$$\Omega[p] = \int_0^\infty \prod_i d\sigma_i P(\sigma) P[D|p, \sigma] \quad (33)$$

One possibility is to assume the  $\delta$ -function prior (30) but with  $\{\sigma_i^{(0)}\}$  fixed and positive, assumed to be known numbers. In this case the integral (33) is trivially calculated with the result  $P[D|p, \sigma^{(0)}]$ . This is in spirit of the frequentist approach to statistics where the noise variance is a fixed albeit unknown number which can be estimated from the data.

Another possibility is to choose a smooth continuous prior for the noise variance. If no information about the noise variance is available, one choice would be to introduce a completely non-informative prior  $P(\sigma) = \text{const}$ . As was noticed by Jaynes (see [1]) this choice is unsatisfactory on the grounds that it does not respect the invariance principle: if we change the variable  $\sigma \rightarrow \sigma'$ , the prior  $P(\sigma)$  will generally change. By requiring that the prior  $P(\sigma)$  should be invariant with respect to a change of scale

$$\sigma^* = a\sigma \quad , \quad P(\sigma^*) = J^{-1}P(\sigma) \quad (34)$$

where in our case the Jacobian  $J = a$ , we obtain the solution

$$P(\sigma) = \frac{\text{const}}{\sigma} \quad (35)$$

This is called Jeffrey's prior. It is of frequent use in Bayesian approaches. Although the prior (35) is unnormalizable (so-called improper prior), its substitution in (33) leads to a converging integral. Another approach considered in the Bayesian literature is to choose prior  $P(x)$  distributions for which a posterior distribution  $P(x|D)$  can be easily calculated. The most popular choice are the so-called conjugate priors [4]. They possess the intuitive feature that starting with a certain functional form for the prior, one ends up with a posterior of the same functional form, but with parameters updated by the sample information. For a Gaussian noise, the conjugate prior is therefore given by the Gamma distribution in  $1/\sigma^{-2}$  :

$$P(\sigma) \propto (\sigma^2)^{1-\alpha} \exp(-\beta\sigma^{-2}) \quad (36)$$

Here  $\alpha > 0$  and  $\beta > 0$ . An improper prior corresponds to the limit  $\alpha \leq 3/2$  and  $\beta \rightarrow 0$ . Note also that the particular choice  $\alpha = \nu/2$ ,  $\beta = 1/2$  corresponds to the Chi-square distribution with  $\nu$  degrees of freedom <sup>6</sup>.

---

<sup>6</sup>This may hint on a bridge to the classical approach to statistics whose fundamental theorem states that if  $S^2$  is the variance of a random sample of size  $n$  from a normal distribution with the mean  $\mu$  and variance  $\sigma^2$ , then the random variable  $(n-1)S^2/\sigma^2$  has a Chi-square distribution with  $(n-1)$  degrees of freedom.

In what follows we analyze both  $\delta$ -function (30) and conjugate (36) choices for the prior. Jeffrey's prior is recovered by the particular choice

$$\alpha = \frac{3}{2} \quad , \quad \beta = 0 \quad (37)$$

in the conjugate prior (36). (As will be seen in the next section, this choice may however be undesirable as it does not lead to a convex optimization problem.) We also note that a conjugate prior  $p(x) \propto x^{\alpha-1} e^{-\beta x}$  can be viewed as the MCE prior maximizing entropy given integral constraints  $E[X] = g_1$ ,  $E[\log(X)] = g_2$ . Recall that the entropy itself can be interpreted as an expected value  $H(X) = E[-\log p(X)]$ , so that the second constraint here actually fixes the entropy of the distribution.

## 6 Effective Lagrangian

If we choose a conjugate prior (36), integral (33) is calculated using the well-known formula for the Gamma distribution

$$\int_0^\infty dy y^{b-1} e^{-ay} = \Gamma(b) a^{-b} \quad (38)$$

and assuming identical distributions for all  $\sigma_i$ . It is convenient here to define the “effective potential”

$$\Phi[p] = -\log \Omega[p] \quad (39)$$

For the conjugate prior (36) we thus obtain

$$\Phi_{CP}[p] = (\alpha - 1) \sum_{i=1}^N \log \left[ \left( \int \int dx dy p(x) p(y|x) f_i(x, y) - C_i \right)^2 + 2\beta \right] \quad (40)$$

plus an inessential constant. If instead the  $\delta$ -function prior (30) is used, the effective potential is (hereafter we omit the subscript (0) from  $\sigma_i^{(0)}$ )

$$\Phi_{DF}[p] = \sum_i \frac{1}{2\sigma_i^2} \left( \int \int dx dy p(x) p(y|x) f_i(x, y) - C_i \right)^2 \quad (41)$$

Using (41), (40), (24) and (29), we find that the MAP probability density  $p(y|x)$  is

$$p(y|x) = \arg \min_{p(y|x)} \max_{\xi(x)} L_{eff}[p, \xi] \quad (42)$$

with the following “effective Lagrangian”:

$$\begin{aligned} L_{eff}[p, \xi] &= h \int dx p(x) \int dy p(y|x) \log \left( \frac{p(y|x)}{q(y|x)} \right) \\ &\quad - \int dx \xi(x) p(x) \left( \int dy p(y|x) - 1 \right) + \Phi[p] \end{aligned} \quad (43)$$

where  $\Phi[p]$  is defined by either (40) or (41).

To find whether the resulting Lagrangian admits a unique minimum or multiple local minima, we study its convexity properties. If the effective Lagrangian (43) is convex, a solution of the minimization problem, if exists, is unique. Recall that a functional  $F[f(x)]$  is convex if

$$F[\lambda f_1(x) + \bar{\lambda} f_2(x)] \leq \lambda F[f_1(x)] + \bar{\lambda} F[f_2(x)] \quad (44)$$

for any distributions  $f_1(x), f_2(x)$  and  $0 \leq \lambda \leq 1$ ,  $\bar{\lambda} \equiv 1 - \lambda$ .

Convexity of the first term in (43) is easy to establish:

$$\begin{aligned} & D[\lambda p_1(y|x) + \bar{\lambda} p_2(y|x) || q] - \lambda D[p_1(y|x) || q] - \bar{\lambda} D[p_2(y|x) || q] \\ &= \lambda \int \int dx dy p(x) p_1(y|x) \log \left( \frac{\lambda p_1(y|x) + \bar{\lambda} p_2(y|x)}{p_1(y|x)} \right) \\ &+ \bar{\lambda} \int \int dx dy p(x) p_2(y|x) \log \left( \frac{\lambda p_1(y|x) + \bar{\lambda} p_2(y|x)}{p_2(y|x)} \right) \leq 0 \end{aligned} \quad (45)$$

because both terms are non-positive by Jensen's inequality (3).

Consider now the effective potential term in (43). Convexity is easy to check for the  $\delta$ -function prior (41) for which

$$\begin{aligned} & \Phi_{DF}[\lambda p_1 + \bar{\lambda} p_2] - \lambda \Phi_{DF}[p_1] - \bar{\lambda} \Phi_{DF}[p_2] \\ &= -\frac{\lambda \bar{\lambda}}{\sigma_0^2} \left[ \int \int dx p(x) (p_1(y|x) - p_2(y|x)) f_i(x, y) \right]^2 \leq 0 \end{aligned} \quad (46)$$

Consider now the effective potential obtained with the conjugate prior (40). Since  $-\log(y)$  is a convex function, convexity of the effective potential (40) is guaranteed if  $0 < \alpha < 1$  and  $\beta = 0$ . For other values of parameters  $\alpha, \beta$ , convexity is lost. In particular, the choice of Jeffrey's prior (37) does not lead to a convex optimization problem. This does not necessarily mean that it cannot give sensible results, but this case is bound to be more complicated than a convex problem, and will not be analyzed here. In what follows, the conjugated prior (36) will therefore be considered only in the parameter range  $0 < \alpha < 1$ ,  $\beta = 0$ .

## 7 Minimization of Effective Lagrangian

As it stands, the effective Lagrangian (43) appears to be considerably more complicated than the corresponding Lagrangian (13) of the canonical MCE method applicable when no noise is present. The difference is due to a non-linearity of the effective potential in (43), which is in contrast to a corresponding linear term in (13). The fact that this potential term is linear enables an analytical solution (15) of the variational problem in the classical MCE method, while a numerical scheme is needed only to fix the values of the Lagrange multipliers. On the contrary, in the present method there are no independent Lagrange multipliers, but the problem seems to boil down to a highly non-linear integral equation due to a non-linearity of the effective potential.

With the MAP approach, a simple trick may be suggested that leads to a drastic simplification of the problem. The trick is akin to one used to treat non-linearities in

quantum field theory. We propose to linearize the effective potential in (43) by introducing new variables  $U_i, \lambda_i$  and minimizing the following Lagrangian<sup>7</sup>

$$L_{ext}[p, \xi, U, \lambda] = \int dx p(x) \int dy p(y|x) \log \frac{p(y|x)}{q(y|x)} - \int dx \xi(x) p(x) \left( \int dy p(y|x) - 1 \right) + \Phi[U] + \sum_i \lambda_i \left( U_i - \int \int dx dy p(x) p(y|x) f_i(x, y) + C_i \right) \quad (47)$$

where  $\Phi[U]$  stands for the effective potential (41) or (40) in which all combinations  $\int \int dx dy p(x) p(y|x) f_i(x, y) - C_i$  are substituted by  $U_i$ :

$$\Phi_{DF}[U] = \sum_i \frac{1}{2h\sigma_i^2} U_i^2, \quad \Phi_{CP}[U] = -\frac{1-\alpha}{h} \sum_{i=1}^N \log [U_i^2] \quad (48)$$

It is easy to see that we come back to the original effective Lagrangian (43) if we first minimize (47) over  $U_i$  and  $\lambda_i$  with  $p(y|x), \xi$  fixed and substitute the result back into (47). The MAP solution  $p(y|x)$  is therefore determined by the following set of minimization equations:

$$\frac{\delta L_{ext}}{\delta p(y|x)} = 0, \quad \frac{\delta L_{ext}}{\delta \xi(x)} = 0, \quad \frac{\partial L_{ext}}{\partial U_i} = 0, \quad \frac{\partial L_{ext}}{\partial \lambda_i} = 0 \quad (49)$$

First two equations in (49) yield

$$p_\lambda(y|x) = \frac{q(y|x)}{Z_\lambda(x)} \exp \left( \sum_i \lambda_i f_i(x, y) \right) \quad (50)$$

where

$$Z_\lambda(x) = \int dy q(y|x) \exp \left( \sum_i \lambda_i f_i(x, y) \right) \quad (51)$$

Up to this point the solution is identical to one given by the canonical MCE method. The difference arises when it comes to a solution for the Lagrange multipliers  $\lambda_i$ . Using (50) and the last two of equations (49), we end up with a set of non-linear equations for  $\lambda_i$ :

$$\int \int dx dy p(x) \frac{q(y|x)}{Z_\lambda(x)} \exp \left( \sum_i \lambda_i f_i(x, y) \right) f_i(x, y) = C_i + U_i^*(\lambda_i), \quad i = 1, \dots, N \quad (52)$$

where  $U_i^*(\lambda_i)$  stands for a solution of the equation

$$\frac{\partial \Phi(U)}{\partial U_i} = -\lambda_i \quad (53)$$

Equation (53) yields

$$\begin{aligned} U_i^*(\lambda_i) &= -h\sigma_i^2 \lambda_i & (DF) \\ U_i^*(\lambda_i) &= \frac{2(1-\alpha)}{h\lambda_i} & (CP) \end{aligned} \quad (54)$$

---

<sup>7</sup>Here we rescale all terms by  $1/h$ .

for the effective potentials (48), respectively. As one can see, for the choice  $\Phi(U) = \Phi_{DF}(U)$  the solution smoothly goes to the canonical MCE solution in the limit  $\sigma_i \rightarrow 0$ .

Let us now analyze convexity properties of the resulting problem of optimization in parameters  $\{\lambda_i\}$ . By the back substitution of the MAP distribution (50) and solution (54) of (53) into (47), we arrive at the following function

$$F(\lambda) \equiv \min_{p(y|x), \xi, U} L_{ext}[p, \xi, U, \lambda] = - \int dx p(x) \log Z_\lambda(x) + \Psi(\lambda) \quad (55)$$

where

$$\Psi(\lambda) = \Phi(U^\star) + \sum_i \lambda_i (U_i^\star + C_i) \quad (56)$$

and  $U_i^\star(\lambda_i)$  is determined by Eq.(53). In particular, for the  $\delta$ -function and conjugate priors we obtain

$$\begin{aligned} \Psi_{DF}(\lambda) &= \sum_i \left( \lambda_i C_i - \frac{h\sigma_i^2}{2} \lambda_i^2 \right), \quad \frac{\partial^2 \Psi_{DF}}{\partial \lambda_i \partial \lambda_j} \leq 0 \\ \Psi_{CP}(\lambda) &= \frac{1-\alpha}{h} \sum_i \log \lambda_i^2 + \sum_i \lambda_i C_i + const, \quad \frac{\partial^2 \Psi_{CP}}{\partial \lambda_i \partial \lambda_j} \leq 0 \end{aligned} \quad (57)$$

For the Hessian matrix we find

$$\begin{aligned} \frac{\partial^2 F}{\partial \lambda_i \partial \lambda_j} &= - \int dx p(x) \left[ \int dy p_\lambda(y|x) f_i(y) f_j(y) - \int dy p_\lambda(y|x) f_i(y) \int dy' p_\lambda(y'|x) f_j(y') \right] \\ &+ \frac{\partial^2 \Psi}{\partial \lambda_i \partial \lambda_j} \equiv - \int dx p(x) cov(f_i, f_j)(x) + \frac{\partial^2 \Psi}{\partial \lambda_i \partial \lambda_j} \leq 0 \end{aligned} \quad (58)$$

which is because a covariance matrix of constraints is positive definite for linearly independent constraints. We conclude that the function  $F(\lambda)$  (55) is concave for both choices of the effective potential, and therefore the solution of the linearized problem (47), if exists, is unique. Furthermore, we note that the derivative of the first term in (55) with respect to  $\lambda_i$  is always non-positive for any value of  $\lambda_i$ . Therefore, a necessary condition that a maximum of function  $F(\lambda)$  is attained at a finite value of  $\lambda_i$  is that the derivative of the second term in (55) should be non-negative. This imposes a restriction on the range of possible values of  $\lambda_i$ :

$$\begin{aligned} \lambda_i &\leq C_i / \sigma_0^2 \quad (DF) \\ \lambda_i &\leq -\frac{2(1-\alpha)}{C_i} \quad (CP) \end{aligned} \quad (59)$$

We thus have reduced the problem of minimization of the effective Lagrangian (43) to a much simpler finite dimensional optimization problem (52)-(54) for the Lagrange multipliers  $\{\lambda_i\}$ , that may be solved by a multidimensional Newton-Raphson method. The computational cost of the method is therefore the same as that of the canonical MCE approach.

## 8 Adding Intermediate Time Levels

The last stage of our construction is a calculation of conditional transition densities involving additional time levels between the shortest and the longest maturities of European options in the calibrating set. In this section we will show how this problem can be consistently addressed using the MCE approach.

In what follows, we will concentrate on the case when a single time level  $T$  is added between two adjacent maturity dates ( $T_k < T < T_{k+1}$ ) in the calibrating set. The distributions  $p(x)$ ,  $p(z|x)$ ,  $p(z)$  of asset's prices  $X$  and  $Z$  at times  $T_k$  and  $T_{k+1}$ , respectively, are supposed to be calculated from the data as described in the previous sections. We will also need prior densities  $q(y|x)$ ,  $q(z|y)$ ,  $q(z|x)$  which may be estimated from the  $p$  probabilities found previously by shifting time arguments as described above. We want to calculate the distribution  $p(y|x) \equiv p(y, T|x, T_k)$ ,  $p(z|y) \equiv p(z, T_{k+1}|y, T)$ . Having solved this problem, we could continue with the same method recursively and fill the model for all additional time levels of interest.

One simple-minded approach to the problem would be to say that as soon as we know that  $p(y|x) \rightarrow \delta(y-x)$  as  $T \rightarrow T_k$  and  $p(y|x) \rightarrow p(z|x)$  as  $T \rightarrow T_{k+1}$ , we could use linear interpolation to estimate transition densities for intermediate times:

$$\begin{aligned}\tilde{p}(y|x) &= (1-\rho) N_y(x, \rho) + \rho p(z|x)|_{z=y} \\ \tilde{p}(z|y) &= (1-\rho) p(z|x)|_{x=y} + \rho N_z(y, 1-\rho)\end{aligned}\tag{60}$$

where  $\rho = (T - T_k)/(T_{k+1} - T_k)$  and  $N_x(m, \sigma^2)$  stands for the normal distribution in  $x$  with mean  $m$  and variance  $\sigma^2$ . A problem with this approach is that such approximate transition densities would generally violate the Markov property

$$p(x, y, z) = p(x)p(y|x)p(z|y)\tag{61}$$

and therefore would not satisfy the Chapman-Kolmogorov equation

$$\int dy p(y|x)p(z|y) = p(z|x)\tag{62}$$

and the consistency equation

$$\int dy p(y)p(z|y) = p(z)\tag{63}$$

Instead of trying to interpolate, we want to approach the problem using the Minimum Cross Entropy approach. This amounts to insisting that the asset price distribution should possess a minimum possible relative entropy with a reference density for arbitrary times, not only for maturity dates of options in the calibrating set. Within this framework, equation (62) can be imposed as a constraint on the MCE solution. The joint KL entropy that should be minimized is

$$D[p(x, y, z)||q(x, y, z)] = D[p(x)||q(x)] + D[p(y|x)||q(y|x)] + D[p(z|y)||q(z|y)]\tag{64}$$

where we used the chain rule and the Markov property. To find the MCE distributions  $p(y|x)$  and  $p(z|y)$ , the entropy (64) should be minimized with the Chapman-Kolmogorov

equation (62) enforced as a constraint, and distributions  $p(x)$  and  $p(z|x)$  fixed. Writing the Lagrange multiplier as  $p(x)\lambda(x, z)$  (which we may do as  $p(x)$  is fixed), we arrive at the Lagrangian

$$\begin{aligned} L &= D[p(x, y, z)||q(x, y, z)] - \int dx dz p(x)\lambda(x, z) \left( \int dy p(y|x)p(z|y) - p(z|x) \right) \\ &= \int dx dy dz p(x, y, z) \log \frac{p(x, y, z)}{q(x, y, z)} - \int dx dz \lambda(x, z) \left( \int dy p(x, y, z) - p(x, z) \right) \end{aligned} \quad (65)$$

Here we used the Markov property (61) to proceed to the second expression. Minimizing this Lagrangian with respect to  $p(x, y, z)$ , we obtain

$$p(x, y, z) = \frac{q(x, y, z)}{q(x, z)} p(x, z) \quad (66)$$

Note that it is crucial for our method that the density  $q$  is non-uniform (at least) in the  $y$  direction, as otherwise equation (66) would indicate that  $p(x, y, z)$  is uniform in  $y$ . This is reasonable because in such a case there would be no either prior information (via a reference density), or constraints on a distribution of  $Y$ . The present method thus adjusts the prior distributions  $q(y|x)$  and  $q(z|y)$  by enforcing the correct joint distribution of the end-point values  $p(x, z)$ .

Using the Markov property of the reference distribution, we find the resulting conditional densities

$$p(y|x) = q(y|x) \int dz \frac{q(z|y)}{q(z|x)} p(x, z) \left[ \int dy dz \frac{q(y|x)q(z|y)}{q(z|x)} p(x, z) \right]^{-1} \quad (67)$$

$$p(z|y) = q(z|y) \int dx \frac{q(y|x)}{q(z|x)} p(x, z) \left[ \int dx dz \frac{q(y|x)q(z|y)}{q(z|x)} p(x, z) \right]^{-1} \quad (68)$$

Equations (67), (68) solve the problem of finding the minimum cross entropy conditional densities  $p(y, T|x, T_k)$  and  $p(z, T_{k+1}|y, T; x, T_k)$ , given reference densities  $q(y|x)$ ,  $q(z|y)$ ,  $q(z|x)$  and fixed transition density  $p(x, z)$ . If necessary, the same procedure can now be continued to get a more detailed partition of the interval  $[T_k, T_{k+1}]$  using the same approach.

## 9 Numerical examples

To illustrate the performance of our method, we generate an artificial “data” with a Monte-Carlo simulation. We will only consider a calculation at one time step of our scheme, and moreover use only the DF prior. A more detailed numerical analysis of our algorithm will be given elsewhere. We take  $x_0 = 3.0$  to be today’s ( $t = 0$ ) asset price. We consider  $N = 5$  call options with strikes

$$K = [2.0, 2.5, 3.0, 3.5, 4.0]$$

Options mature at  $T = 1$  year. Risk-free interest rate is  $r = 0.1$ . We assume that the “true”, noiseless prices are given by the Black-Scholes model with volatility  $\sigma_{BS} = 0.25$ . This yields the following prices

$$C = [1.19, 0.78, 0.45, 0.23, 0.11]$$

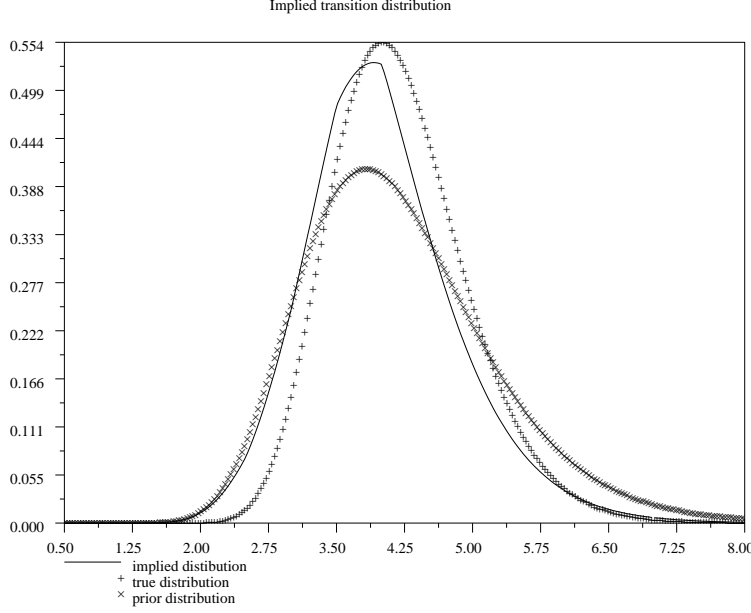


Figure 1: Implied transition density calculated with the DF prior.  $\sigma\sqrt{h} = 0.45, x = 4.0$ .

We model the market noisy prices  $\{\hat{C}_i\}$  as follows:

$$\hat{C}_i = C_i [1 + \sigma_{ns} N(0, 1)] \quad (69)$$

where  $\sigma_{ns}$  measures the noise in the data, and  $N(0, 1)$  is a random number drawn from a normal distribution with zero mean and unit variance. In our simulation we have chosen  $\sigma_{ns} = 0.15$  (i.e. 15 % noise in prices). We have considered the following set of “market prices” obtained with such simulation:

$$\hat{C} = [1.07, 0.68, 0.53, 0.22, 0.10]$$

Using this data as an input, we want to calculate with our method the transition density  $p(y, T|x, t)$  for  $t = 0.5$ . For the cumulative density  $p(x)$  at time  $t = 0.5$ , we will assume the lognormal distribution with the same “true” value  $\sigma_{BS} = 0.25$ , while for the prior transition density  $q(y|x)$  we take the lognormal distribution with  $\sigma_{BS} = 0.35$ . For the effective parameter  $\sigma\sqrt{h}$  controlling the relative weights of the entropy and likelihood terms, we have found that the interval  $\sigma\sqrt{h} = [0.45, 3.0]$  provide a “stability plateau” between overfitting (for lower values of  $\sigma\sqrt{h}$ ) and oversmoothing (for higher values of  $\sigma\sqrt{h}$ ). Results for  $\sigma\sqrt{h} = .45$  and  $\sigma\sqrt{h} = 3.0$  are shown on Figs.1 to 2 and Figs. 3 to 4, respectively. Figs. 5 and 6 illustrate the appearance of a fake binodal structure in the implied densities at a low value  $\sigma\sqrt{h} = 0.1$ . For lower values of  $\sigma\sqrt{h}$ , the algorithm blows up, in agreement with results by Buchen and Kelly [6]. This shows directly that juggling of an implied distribution in the canonical MCE method is due to overfitting.

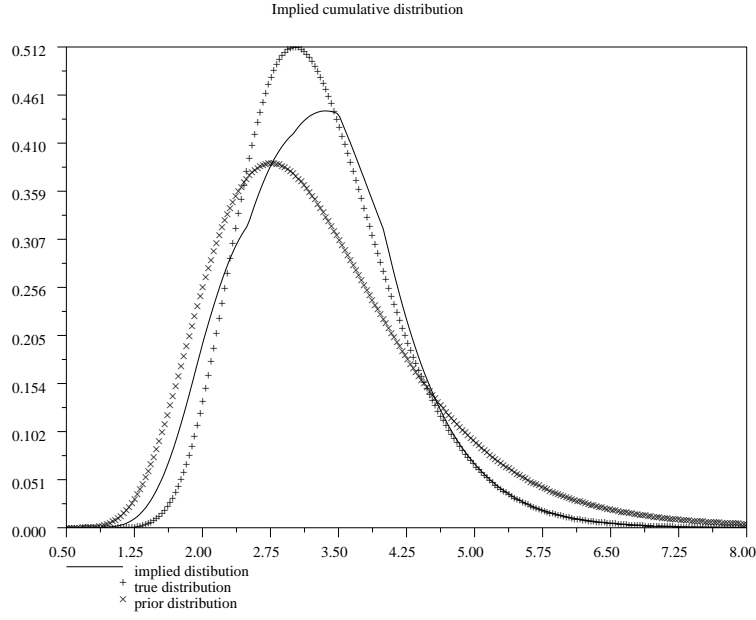


Figure 2: Implied cumulative density calculated with the DF prior.  $\sigma\sqrt{h} = 0.45$ .

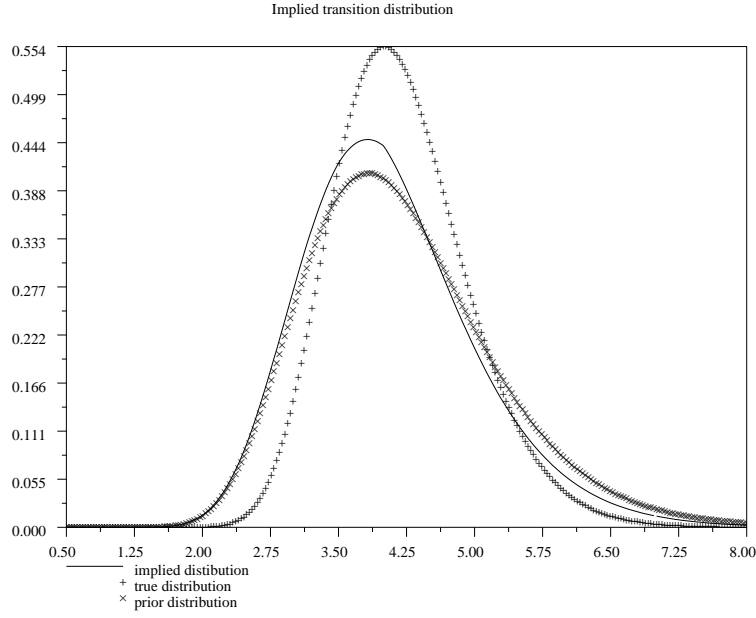


Figure 3: Implied transition density calculated with the DF prior.  $\sigma\sqrt{h} = 3.0, x = 4.0$ .

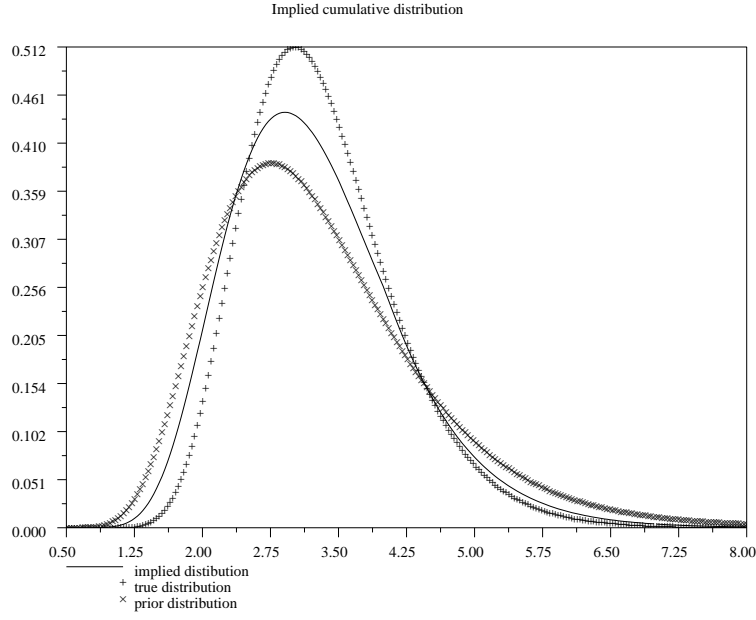


Figure 4: Implied cumulative density calculated with the DF prior.  $\sigma\sqrt{h} = 3.0$ .

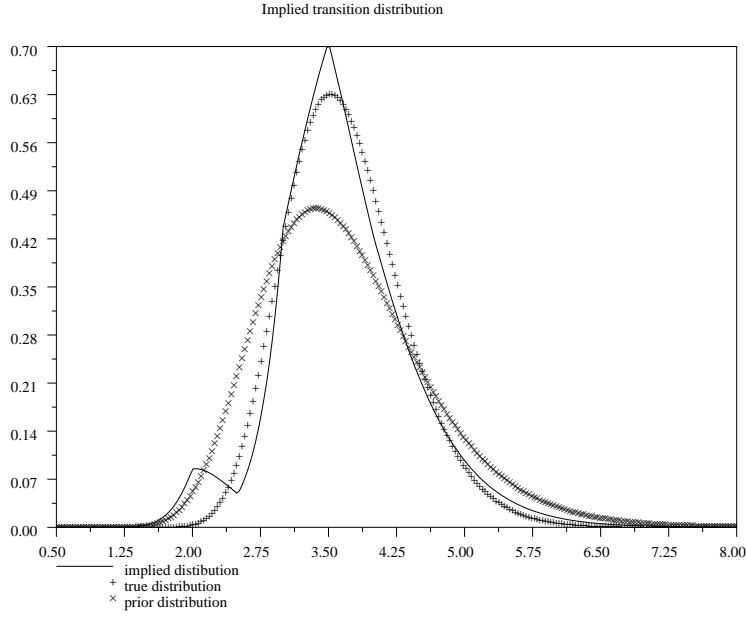


Figure 5: Implied transition density calculated with the DF prior.  $\sigma\sqrt{h} = 0.1, x = 3.5$ .

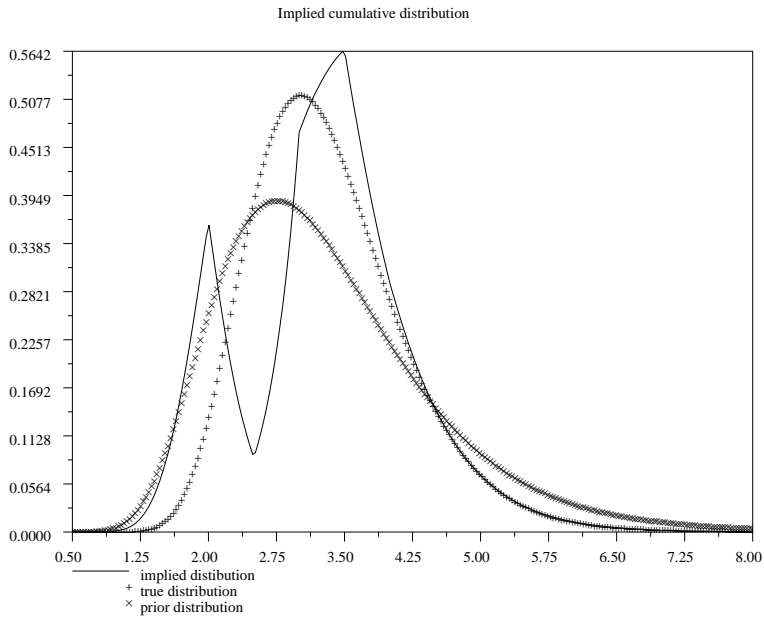


Figure 6: Implied cumulative density calculated with the DF prior.  $\sigma\sqrt{h} = 0.1$ .

## References

- [1] E.T. Jaynes, *Probability Theory: The Logic of Science*, (1994), <http://Bayes.Wustl.Edu>.
- [2] G.L. Bretthorst, “ An Introduction to Parameter Estimation Using Bayesian Probability Theory”, in *Maximum Entropy and Bayesian Methods*, P.F. Fougere (ed.), Kluwer Academic Publishers (1990), <http://Bayes.Wustl.Edu>.
- [3] *Maximum Entropy and Bayesian Methods*, J. Skilling (ed.), Kluwer Academic Publishers (1989).
- [4] A. Gelman, J.B. Carlin, H.S. Stern and D.R. Rubin, *Bayesian Data Analysis*, Chapman & Hall (1995).
- [5] L. Gulko, “The Entropy Theory of Option Pricing” *Yale University Working Paper* (1995); “The Entropy Pricing Theory”, *Ph.D. Thesis, Yale School of Management* (1998).
- [6] P.W. Buchen and M.F. Kelly, “The Maximum Entropy Distribution of an Asset Inferred from Option Prices”, *Journal of Financial and Quantitative Analysis*, **31** (1996) 143.
- [7] M. Stutzer, “A Simple Nonparametric Approach to Derivative Security Valuation”, *Journal of Finance*, **51** (1996) 1633.
- [8] T.M. Cover and J.L. Thomas, *Elements of Information Theory*, Wiley, 1991.
- [9] M. Avellaneda, R. Holmes. C. Friedman and D. Samperi, “Calibration of Volatility Surfaces via Relative-Entropy Minimization”, *Applied Mathematical Finance*, March 1997, 4;  
M. Avellaneda, “Minimum Relative Entropy Calibration of Asset-Pricing Models”, *International Journal of Theoretical and Applied Finance* **1** (1998), 447.
- [10] E. Derman and J. Zou, “A Fair Value for the Skew”, *RISK Magazine*, January 2001, 111.
- [11] M. Stutzer, “Simple Entropic Derivation of Generalized Black-Scholes Option Pricing Model”, *Entropy* **2** (2000) 70, <http://www.mdpi.org/entropy/>.
- [12] E.T. Jaynes, “ On the Rationale of Maximum-Entropy Methods”, *Proceedings of the IEEE*, **70**, No. 9 (1982), 939, <http://Bayes.Wustl.Edu>.
- [13] P.W. Buchen and M.F. Kelly, “The Maximum Entropy Distribution of an Asset Inferred from Linear Inverse Theory”, *Journal of Computational Finance*, **3** (2000) 53.
- [14] M. Rubinstein, “Implied Binomial Trees”, *Journal of Finance*, **49** (1994) 771.
- [15] E. Derman and I. Kani, “Riding on a Smile”, *RISK Magazine*, **7**, No. 2, (1994) 32.

- [16] J.C. Jackwerth, “Option-Implied Risk-Neutral Distributions and Implied Binomial Trees: A Literature Review”, *Journal of Derivatives*, **7**, No.2 (1999) 66.
- [17] G. Brown and K.B. Toft, “Constructing Binominal Tree from Multiple Implied Probability Distributions”, *Journal of Derivatives*, **7**, No.2 (1999) 83.
- [18] N. Agmon, Y. Alhassid and R.D. Levine, “ An Algorithm for Determining the Lagrange parameters in the Maximum Entropy Formalizm”, in *The Maximum Entropy Formalizm*, Ed. R.D. Levine and M. Tribus, Cambridge MA: MIT Press (1981), 207-209.
- [19] C.C. Rodriguez, “Entropic priors” (1991), <http://omega.albany.edu:8008>.
- [20] A. Caticha, “Maximum Entropy, Fluctuations and Priors”, presented at MaxEnt 2000, <http://xxx.lanl.gov/abs/math-ph/0008017>.